

I. G. Zenkevich

DOES REFERENCE DATA ON PHYSICOCHEMICAL PROPERTIES OF ORGANIC COMPOUNDS NEED AN ADDITIONAL PROCESSING?

St. Petersburg State University,

7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

Using normal boiling points of *n*-alkyl phosphines it is shown that some of available reference data on the physicochemical properties of organic compounds are unreliable or even erroneous. The use of such data for any practical purposes is undesirable. Among known methods for evaluating physicochemical characteristics of single compounds there are no ones providing a priori required precision. However, such demand can be fulfilled using the methods of comparison of properties of different series of homologues. The predestination of the quality control approach considered is refining physicochemical properties (*A*) of homologues. It implies their comparison (mutual recalculation) with analogous data using three-parameter linear-logarithmic equation $\lg A(\text{series I}) = a \lg A(\text{series II}) + bY + c$, where *A* is the values of the property under consideration, *Y* is the additive parameter providing the mutually unambiguous correspondence of homologs of two series under comparison. In this paper the condition $Y = nC$ (number of carbon atoms in the molecules of homologs under consideration) is selected. The following independent step of data processing is the controlling of correspondence of *A*-values to the recurrent relation $A(n+1) = aA(n) + b$, where $A(n+1)$ is the *A*-value for homolog containing (*n* + 1) carbon atoms in the molecule, and $A(n)$ is the *A*-value for previous homolog. Similar procedures of the quality control of reference data is recommended for different properties of organic compounds. Refs 21. Figs 1. Tables 1.

Keywords: organic compounds, physicochemical characteristics, quality control of reference data, *n*-alkyl phosphines, normal boiling points.

И. Г. Зенкевич

НУЖДАЮТСЯ ЛИ СПРАВОЧНЫЕ ЗНАЧЕНИЯ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ В ДОПОЛНИТЕЛЬНОЙ ОБРАБОТКЕ?

Санкт-Петербургский государственный университет,

Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

На примере нормальных температур кипения *n*-алкилфосфинов показано, что некоторые из известных справочных значений физико-химических свойств органических соединений ненадёжны или даже ошибочны. Практическое использование таких данных нежелательно. Среди известных в настоящее время методов оценки физико-химических характеристик отдельных соединений нет таких, которые обеспечивали бы приемлемую для практических целей заданную точность результатов. Однако такие способы есть среди методов сравнения свойств совокупностей гомологов разных рядов. Предлагаемый алгоритм проверки и, при необходимости, уточнения физико-химических характеристик (*A*) гомологов основан на их сравнении с аналогичными данными для одного или нескольких рядов подробнее охарактеризованных структурных аналогов. Последующая обработка этих данных включает взаимный пересчёт с использованием линейно-логарифмического соотношения $\lg A(\text{ряд I}) = a \lg A(\text{ряд II}) + bY + c$, где *A* — значения рассматриваемого свойства; *Y* — аддитивная величина, обеспечивающая взаимно-однозначное соответствие гомологов сравниваемых друг с другом рядов. Независимой стадией контроля данных является проверка наборов значений *A* на соответствие рекуррентным соотношениям $A(n+1) = aA(n) + b$, где $A(n+1)$ — значение свойства *A* для гомолога, содержащего (*n* + 1) атомов углерода в молекуле; $A(n)$ — значение этого же свойства для предыдущего гомолога. Аналогичные

процедуры контроля справочных значений различных свойств органических соединений рекомендованы для соединений разных рядов. Библиогр. 21 назв. Ил. 1. Табл. 1.

Ключевые слова: органические соединения, физико-химические характеристики, контроль справочных данных, *n*-алкилфосфины, нормальные температуры кипения.

Introduction. In practice of organic chemistry physicochemical characteristics are typically used for checking the purity of compounds and their identification. Their variety includes normal (at the atmospheric pressure) boiling points (T_b), melting points (T_m), relative densities (d_4^{20}), indices of refraction (n_D^{20}), and others (used more rarely). So far as the determination of such properties implies often time-consuming isolation of individual compounds from reaction or natural mixtures, the value of such data significantly reduced at present due to preferable using the spectral methods (at first, NMR and mass spectrometry), those are applicable to mixtures. However, the “classical” physicochemical properties of organic compounds can be recalculated with high precision into gas chromatographic retention indices [1] which are analytical parameters of high importance, because they can be applicable both to the individual compounds and to the constituents of complex mixtures, as well.

This article is devoted to the characterization of the algorithm for testing and clarification of reference data on the example of normal boiling points of rather “exotic” class of organic compounds — *n*-alkyl phosphines, $R-PH_2$ (in accordance with contemporary IUPAC nomenclature these compounds should be named as *n*-alkyl phosphanes). The choice of *n*-alkyl phosphines is explained by clearly expressed unreliability of their reference T_b values, that is caused by high reactivity of these compounds and their easy oxidation (up to the manifestation of pyrophoric properties by few simplest homologs). In the result, the synthesis of these compounds with the aim of more precise determination of their physicochemical properties is considered to be injudicious; for such purposes the theoretical methods are used preferably. However, the accuracy of known methods for precalculation of physicochemical properties including those of ACD (Advanced Chemical Development) software depends on the chemical origin of compounds and for alkyl phosphines it seems to be rather low.

Alkyl phosphines are restrictedly used in organic chemistry, despite of their easy synthetic availability [2–4], possess high toxicity and can be formed in biochemical processes [5]. The level of their characterization with analytical parameters for GC-MS identification is similar to many other “exotic” series. If their standard mass spectra (electron ionization) are well known (in MS database [6] we can find mass spectra of 18 alkyl phosphines C_1-C_{13}), than GC retention indices are not determined nor for one compound of this series. It should be noted that one of the most accurate methods of their evaluation is based on the use of just normal boiling points.

Experimental (initial data and their processing). Reference values of normal boiling points of *n*-alkyl phosphines were taken from various available sources including Beilstein-based reference books (e. g., [7]), original publications (e. g., [8]), and Internet sources of information. So far as the aim of this paper (it does not purport to completeness of the revealing all of them) is testing and clarifying the initial T_b values, the indication of the original references was considered to be irrational.

The calculation of coefficients of three-parameter equation (1) and data processing using this equation by method of least squares was fulfilled using QBasic program, the calculations using recurrent linear regression relation (2) — using Origin software (versions 4.1 and 8.1). For independent evaluations of boiling points was used ACD software (version of years 1994–1996).

Results and their discussion. Available reference values of normal boiling points (T_b) of n -alkyl phosphines $R-PH_2$ (R is C_nH_{2n+1} , where $1 \leq n \leq 10$) are presented in the second column of Table. Besides them T_b value for phenyl phosphine (160–161°C) [8] is known.

**Comparison of available reference values of normal boiling points
of n -alkyl phosphines $C_nH_{2n+1}-PH_2$ ($1 \leq n \leq 10$) and additional
data used for their testing and refining**

R	Reference T_b values for RPH_2 , °C	Calculated values T_b (ACD), °C	Calculated T_b values from data for RSH , °C ^c	Accepted T_b values for RPH_2 , °C	Reference T_b values for RSH , °C
CH ₃	–17.1; –14	–13.2		–15 ± 2^e	6.0
C ₂ H ₅	25	24.5	25^d	25^f	35
C ₃ H ₇	– ^a	57.6	50.2	50	67.6
C ₄ H ₉	54; 60; 76 ^b	87.4	76	76	98.5
C ₅ H ₁₁	101.8; 102	114.8	102.2	102	126.5
C ₆ H ₁₃	127.8	140.2	129.0	128	152
C ₇ H ₁₅	–	164.1	156.8	157	177
C ₈ H ₁₇	168.8; 169; 184–187	186.6	185.5	186	199
C ₉ H ₁₉	186.8; 187	208.0	215.3	215	220
C ₁₀ H ₂₁	–	228.4		–	240

^a Dash corresponds to the absence of data;

^b Using T_b value for isobutyl phosphine (62°C) permits us to exclude the values 54 and 60°C for n -butyl phosphine (see comments in the text);

^c Coefficients of equation $\log T_b(RPH_2) = a \log T_b(RSH) + bn_C + c$ (1) are: $a = 0.3205$, $b = 4.899 \cdot 10^{-2}$, $c = 3.762$, n_C is the equal number of carbon atoms in molecules of homologs from series compared;

^d T_b values selected as reference data for calculation of the coefficients of equation (1) are marked in bold;

^e The result of the averaging of all T_b values for methyl phosphine;

^f T_b values corresponded to some initial reference data are marked in bold.

At first, the highly heterogeneous character of initial data should be underlined. Three from ten homologs (C₃, C₇ and C₁₀) are not characterized by experimental T_b values up to present, as well as T_b values for homologs with $n_C > 10$ remain unknown. For few homologs some auxiliary indirect evaluations are possible. Thus, for n -propyl phosphine we can postulate $T_b > 41^\circ\text{C}$ because its value cannot be less than T_b of isopropyl phosphine (41°C).

Only single T_b values indicated with different accuracy were found for four alkyl phosphines (C₂, C₅, C₆ and C₉). The value 25°C for ethyl phosphine means that its accuracy is approximately equal to the unit of the last digit presented in the number, i. e., $25 \pm 1^\circ\text{C}$, while the accuracy of the values 101.8, 127.8 and 186.8°C is ten times more ($\pm 0.1^\circ\text{C}$). It should be noted that determination of boiling points with accuracy about one decimal units of centigrade is possible for stable organic compound, which contain minimal amounts of impurities (is it the obligatory condition for preventing the formation of azeotropic mixtures). It is unlikely that primary alkyl phosphines meet these requirements; more likely that boiling points presented with decimal units are precalculated values indicated without mention of the calculation method. Such "replacing" of experimental data with precalculated ones seems to be typical for ChemSpider and other Internet databases for properties of organic compounds. A distinctive sign of experimental data is the indication of ranges, as it is for n -octyl phosphine (184–187°C). In accordance with this criterion the T_b values for C₅ and C₆ homologues (101.8 and 127.8°C) can be considered as "suspicious" as well.

Among other incorrect values it should be mentioned the value for nonyl phosphine (186.8°C) that is equal to one of T_b values for octyl phosphine (184–187°C). Hence, at least

one of them is erroneous. Finally, we should mention that three homologs (C_1 , C_4 and C_8) are characterized by few values not matching with each other. It seems to be typical not only for alkyl phosphines, but for compounds of other homologous series, as well. Sometimes the relation to such discrepancies of reference values is determined a priori by the credibility of different authors that is used rather often in physical chemistry (Prof. E. P. Sokolova, St. Petersburg State University, personal communication). However, the author of this article is not shared such discrimination.

Further data processing implies excluding the erroneous data; this step often appears to be rather complex and cannot be algorithmized unambiguously. Thus, methyl phosphine is characterized by two T_b values, namely -17.1 and -14°C . The first (not precalculated) of them seems to be preferable, because it is indicated with higher precision. Three T_b values (54 , 60 and 76°C) are known for *n*-butyl phosphine. In such cases it seems to be reasonable to compare every of these different values with data for isomers of compound under characterization. Boiling point of isobutyl phosphine is 62°C , while that for *tert*-butyl phosphine is 53 – 55°C . Hence, two from known three T_b values for butyl phosphine can be attributed with these isomers, while the rest is the most reliable T_b value for target compound (76°C). Among two T_b values for *n*-octyl phosphine (168.8 and 184 – 187°C) the second one presented in the form of range seems to be preferable.

The algorithm of T_b evaluations used in ACD software is based on the modified additive scheme; hence, its accuracy cannot be high. Such evaluations are presented in the third column of Table. Some of them can be found in the Internet (e. g., ChemSpider). The reasonable coincidences of ACD-precalculated values with experimental data (at least within the ranges $\pm 10^\circ\text{C}$) are observed for three homologs (C_1 , C_2 and C_8) only. Hence, this approach cannot be recommended as reliable criterion for quality control of reference data.

Summarizing these comments, we can notice that the numerous methods of normal boiling points evaluation known at present (see, e. g., [9–17]) there are no those providing the specified accuracy of the results. However such examples we can find among methods of comparison of physicochemical data for sets of homologs of different series. In other words, checking or reconsidering the values of physicochemical properties of single organic compounds is the complex task, while the solution of similar problems for series of homologs appears to be much simpler.

Just the possibility of controlling the accuracy of results is one of the principal advantages of the concept of comparative recalculation of physicochemical properties proposed by M. Kh. Karapetyantz yet in the middle of 1960s [18]. However, the linear-logarithmic equation which permits us to provide such recalculations for normal boiling points was suggested later [19, 20]:

$$\log T_b(\text{series I}) = a \log T_b(\text{series II}) + bY + c, \quad (1)$$

where Y is the additive parameter providing the mutually unambiguous conformity of homologs of series under comparison. The following variables can be used as equivalent: molecular mass (M), the number of carbon or hydrogen atoms in the molecule (n_C , n_H), molar refraction (MR_D), and others; in this paper the selection $Y = n_C$ is used. Coefficients a , b , and c are calculated by method of least squares. All T_b values before calculation of logarithms are recalculated in the scale of absolute temperatures.

The dependencies $T_b(n_C)$ are non-linear for any homologous series, but this non-linearity is well approximated by equation (1). The starting point for application of this approach for any set of homologs is the selection of the homologous series of their structural analogues. Of course, such choice may be extremely wide. For prediction of boiling points of *n*-alkyl

phosphines, R-PH₂, it seems most reasonable to select boiling points of isobaric *n*-alkyl mercaptans, R-SH. Despite of easy oxidation of mercaptans by air oxygen, these compounds are widely used in industry (namely, as odorants of common gas), so T_b values for them are determined with high reliability (see 6th column of Table). The following fact can be used as an auxiliary criterion of correctness of the results: the boiling points of alkyl mercaptans RSH exceed the boiling points of alkyl phosphines R-PH₂ with the same R. This rule is valid not only for R = C_nH_{2n+1}, but for R = phenyl (169–170°C (phenyl mercaptan) vs. 160–161°C (phenyl phosphine)).

So far as the relation (1) is three-parameter equation, the calculation of its coefficients a , b and c requires selection of data not less than for three pairs of equivalent homologs of two series T_b (series II), n_C , T_b (series I). Hence, among the data for alkyl phosphines we should select at least for three most reliable values. Some of residual data can be used for controlling of calculations. Thus, let us select the data for the following homologs: C₂ (25°C), C₄ (76°C) and C₈ (185.5°C, the middle of the range 184–187°C). The coefficients of the equation (1) calculated using these data are presented as the footnote to Table. After that we recalculate the reference T_b data for alkyl mercaptans into T_b values for *n*-alkyl phosphines (indicated in the 4th column of Table).

The results obtained for homologs C₅ and C₆ well correspond to the known reference T_b values for them (101.8 and 127.8°C). It confirms both the reliability of these values and correctness of the approach at whole. However it has the principal limitation: it provides the high precision of results for homologs C₂–C₈ (interpolation), but cannot be recommended for their extrapolation for homologs both with $n_C < 2$ (methyl phosphine) and $n_C > 8$. If the T_b evaluation for C₉-homolog still can be taken into account, because it satisfy the criterion $215 < 220$ (T_b of nonyl mercaptan), but evaluated value for C₁₀-homolog is obviously erroneous ($246 > 240$) and it should be rejected.

The fifth column of Table includes the final normal boiling points which are recommended for using as corrected reference data after their checking and clarification of “raw” reference data. The simplest homolog of this series — methyl phosphine — is characterized by the value $-15 \pm 2^\circ\text{C}$, that is the result of the averaging of all values known for this compound.

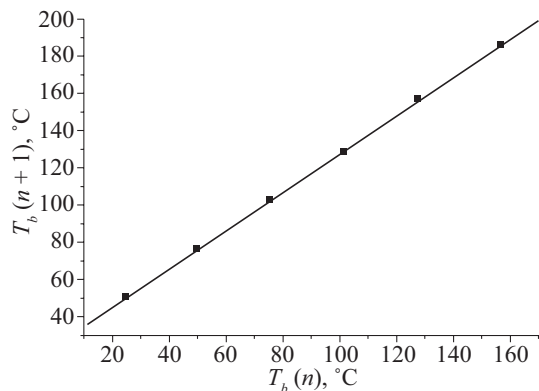
For evaluation of boiling points we can select the data for any homologous series instead of alkyl mercaptans; it does not affect the results. Thus, the selection of data for *n*-alkyl amines C₁–C₁₀ keeping the same reference homologs C₂, C₄ and C₈ we obtain the following T_b evaluations for *n*-alkyl phosphines, °C: 25.0 (C₂), 50.4 (C₃), 76.0 (C₄), 101.8 (C₅), 129.7 (C₆), 156.9 (C₇), 185.5 (C₈), and 215.1 (C₉). Their differences from the data presented in Table are less then 1°C. Obviously, the coefficients of equation (1) in this case are different, namely: $a = 0.3517$, $b = 4.555 \cdot 10^{-2}$, $c = 3.613$. Moreover, the equation (1) allows consolidating the data for normal linear homologs and isomers with branched carbon skeletons; unfortunately of characterized such isomers for alkyl phosphines is equal to only two: R = *iso*-C₃H₇ ($T_b = 41^\circ\text{C}$) and R = *tert*-C₄H₉ (54°C). In the results we obtain another set of values of the coefficients of relation (1): $a = 0.6765$, $b = 2.514 \cdot 10^{-2}$, $c = 1.757$. This extension of the chemical domain of equation (1) provides its new applications, e. g., we can evaluate T_b of 2-heptyl phosphine using T_b of 2-heptyl mercaptan (163.6°C), that is $149.1 \approx 149^\circ\text{C}$.

Despite of numerous verifications of the correctness of the equation (1) for various homologous series [19, 20], it seems reasonable to provide an additional quality control of the results in each partial case. To do that we can use the following regularity: any physico-chemical properties (A) of homologues obey the first order recurrent relations [21]:

$$A(n+1) = aA(n) + b, \quad (2)$$

where $A(n+1)$ is the A -value for homolog having $(n+1)$ carbon atoms in a molecule, $A(n)$ is the value of the same property for previous homolog having n carbon atoms in a molecule, coefficients a and b are calculated by method of least squares. For most cases the correlation coefficients are not less than 0.999.

The plot corresponding to recurrent relation (2) for normal boiling points of n -alkyl phosphines C_1-C_8 is shown on Figure. The parameters of linear regression (2) are presented in the legend to this figure. So far as the value of correlation coefficient is $r = 0.99989$, the set of points at Figure visually well corresponds to the straight line. Hence, all the data from this set can be considered as sufficiently reliable values.



The plot of linear recurrent regression $T_b(n+1) = aT_b(n) + b$ for normal boiling points of n -alkyl phosphines $R-PH_2$ used for additional quality control of data:

Parameters of equation are: $a = 1.03 \pm 0.01$, $b = 24.2 \pm 0.8$, $r = 0.99989$, $S_0 = 0.8$

Thus, reverting to the question used as the title of this article, it should be noted that the procedures of quality control and, if necessary, clarification of reference data on various physicochemical properties of organic compounds are clearly required. Really, this act is rather often used in chemical practice.

All the readers who will find the inconsistencies between the evaluations obtained and new experimental values for normal boiling points of n -alkyl phosphines are requested to inform the author or the journal Editorial office.

References

1. Zenkevich I.G. New methods of precalculation of chromatographic retention indices. Application of linear-logarithmic equations, principles of structural analogy and methods of molecular dynamics. *100 Years of Chromatography. Collected papers*. Moscow, Khimia Publ, 2003. pp. 311–336. (In Russian)
2. Purdela D., Vylchanu R. *Chemistry of organic compounds of phosphorus*. In translation from Romanian to Russian. Moscow, Khimia Publ., 1972. 752 p.
3. Petrov K. A., Parshina V. A. Oxyalkyl phosphines and oxyalkyl phosphinoxides. *Rus. Chem. Rev.*, 1968, vol. 37, no. 1, pp. 1218–1242.
4. Zhang W., Shu M. Reduction of activated carbonyl group by alkyl phosphines: formation of α -hydroxy esters and ketones. *Chem. Commun.*, 2006, pp. 1218–1220.
5. Roels J., Verstraete W. Biological formation of volatile phosphorous compounds, a review paper. *Biores. Technol.*, 2001, vol. 79, no. 3, pp. 243–250.
6. The NIST 14 Mass Spectral Library (NIST14/2014/EPA/NIH). Software/Data Version (NIST 14); NIST Standard Reference Database, Number 69, August 2014. National Institute of Standards and Technology, Gaithersburg, MD 20899. Available at: <http://webbook.nist.gov> (accessed: August 2016).
7. *Properties of organic compounds. Reference book*. Ed. by A. A. Potekhin. Leningrad, Khimia, 1984. 519 p. (In Russian)
8. Beg M. A. A. Correlation of the physical properties of organophosphorus compounds. I. Density, boiling point, and the heat of evaporation of phosphines. *Bull. Chem. Soc. Jap.*, 1967, vol. 40, pp. 15–18.

9. Pearson D. E. A method of estimating the boiling points of organic liquids. *J. Chem. Educ.*, 1961, no. 2, pp. 60–62.
10. Anderson H. H. Boiling points and boiling point numbers of alkyl derivatives of bivalent S, Se, and Te. *J. Chem. Eng. Data*, 1964, vol. 9, no. 2, pp. 272–275.
11. Screttas C. G., Heropoulos G. A. Transferability-additivity of molar volumes of organic liquids and their relation to normal boiling points. *J. Org. Chem.*, 1993, vol. 58, no. 14, pp. 3654–3659.
12. Stein S. E., Brown R. L. Estimation of normal boiling points from group contributions. *J. Chem. Inf. Comput. Sci.*, 1994, vol. 34, no. 3, pp. 581–587.
13. Katritzky A. R., Mu L., Lobanov V. S., Karelson M. Correlation of boiling points with molecular structures. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.*, 1996, vol. 100, pp. 10400–10407.
14. Rucker G., Rucker C. On topological indices, boiling points, and cycloalkanes. *J. Chem. Inf. Comput. Sci.*, 1999, vol. 39, pp. 788–802.
15. Tomovic Z., Gutman I. Modeling boiling points of cycloalkanes by means of iterated line graph sequences. *J. Chem. Inf. Comput. Sci.*, 2001, vol. 41, pp. 1041–1045.
16. Lim T.-C. Mass-modified Wiener indices and the boiling points for lower chloroalkanes. *Acta Chim. Slov.*, 2004, vol. 51, pp. 611–618.
17. Gonzalez M. P., Toropov A. A., Duchowicz P. R., Castro E. A. QSPR calculation of normal boiling points of organic molecules based on the use of correlation weighting of atomic orbitals with extended connectivity of zero- and first-order graphs of atomic orbitals. *Molecules*, 2004, vol. 9, pp. 1019–1033.
18. Karapetyantz M. Kh. *Methods of comparative calculation of physicochemical properties*. Moscow, Khimia Publ., 1965. 401 p. (In Russian)
19. Zenkevich I. G. Evaluation of boiling points of organic compounds using their mutual correlation within different taxonomic groups. *Res. J. Phys. Chem.*, 1996, vol. 70, no. 1, pp. 33–38.
20. Zenkevich I. G. New approach to the equations for evaluation of boiling points of homologs using their positions within corresponding homologous series. *Rus. J. Org. Chem.*, 1998, vol. 34, no. 10, pp. 1458–1462.
21. Zenkevich I. G. Recurrent relations for approximation of physicochemical properties of homologs. *Rus. J. Phys. Chem.*, 2008, vol. 82, no. 5, pp. 695–703.

Статья поступила в редакцию 29 августа 2016 г.

Контактная информация

Zenkevich Igor Georgievich — Doctor of Chemistry, Professor; e-mail: izenkevich@mail15.com

Зенкевич Игорь Георгиевич — доктор химических наук, профессор; e-mail: izenkevich@mail15.com